Application of Information Theory, Lecture 1
Basic Definitions and Facts

Iftach Haitner

Tel Aviv University.

March 08, 2018

The entropy function

X — Discrete random variable (finite number of values) over $\mathcal{X}$ with probability mass $p = p_X$.

The entropy function

X — Discrete random variable (finite number of values) over $\mathcal{X}$ with probability mass $p = p_X$. The entropy of X is defined by:

$$H(X) := -\sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log_2 \Pr[X = x]$$

taking $0 \cdot \log 0 = 0$.

The entropy function

X — Discrete random variable (finite number of values) over $\mathcal{X}$ with probability mass $p = p_X$. The entropy of X is defined by:

$$H(X) := -\sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log_2 \Pr[X = x]$$

taking $0 \cdot \log 0 = 0$.

- $H(X) = -\sum_x p(x) \log p(x) = \mathrm{E}_X \log \frac{1}{p(X)} = \mathrm{E}_{Y=p(X)} \log \frac{1}{Y}$

The entropy function

X — Discrete random variable (finite number of values) over $\mathcal{X}$ with probability mass $p = p_X$. The entropy of X is defined by:

$$H(X) := -\sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log_2 \Pr[X = x]$$

taking $0 \cdot \log 0 = 0$.

- $H(X) = -\sum_x p(x) \log p(x) = \mathrm{E}_X \log \frac{1}{p(X)} = \mathrm{E}_{Y=p(X)} \log \frac{1}{Y}$
- $H(X)$ was introduced by Shannon as mesure for the uncertainty in X — number of bits requited to describe X, information we don't have about X.

The entropy function

X — Discrete random variable (finite number of values) over $\mathcal{X}$ with probability mass $p = p_X$. The entropy of X is defined by:

$$H(X) := -\sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log_2 \Pr[X = x]$$

taking $0 \cdot \log 0 = 0$.

- $H(X) = -\sum_x p(x) \log p(x) = \mathrm{E}_X \log \frac{1}{p(X)} = \mathrm{E}_{Y=p(X)} \log \frac{1}{Y}$
- $H(X)$ was introduced by Shannon as mesure for the uncertainty in X — number of bits requited to describe X, information we don't have about X.
- When using the natural logarithm, the quantity is called nats (“natural")

The entropy function

X — Discrete random variable (finite number of values) over $\mathcal{X}$ with probability mass $p = p_X$. The entropy of X is defined by:

$$H(X) := -\sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log_2 \Pr[X = x]$$

taking $0 \cdot \log 0 = 0$.

- $H(X) = -\sum_x p(x) \log p(x) = \mathrm{E}_X \log \frac{1}{p(X)} = \mathrm{E}_{Y=p(X)} \log \frac{1}{Y}$
- $H(X)$ was introduced by Shannon as mesure for the uncertainty in X — number of bits requited to describe X, information we don't have about X.
- When using the natural logarithm, the quantity is called nats (“natural")
- Entropy is a function of p (sometimes refers to as $H(p)$).

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0, 1\}^n$:

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathsf{P}_X(x_1) = \frac{1}{2}, \mathsf{P}_X(x_2) = \frac{1}{4}, \mathsf{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathsf{P}_X(x_1) = \frac{1}{2}, \mathsf{P}_X(x_2) = \frac{1}{4}, \mathsf{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with
   $\mathsf{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$.

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with
   $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$.

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with
   $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) = ?$

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with
   $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) = ?$
5. $X \sim (p, q)$, $p + q = 1$

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) = ?$
5. $X \sim (p, q)$, $p + q = 1$

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) = ?$
5. $X \sim (p, q)$, $p + q = 1$

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) = ?$
5. $X \sim (p, q)$, $p + q = 1$

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) =?$
5. $X \sim (p, q)$, $p + q = 1$
   - $H(X) = H(p,q) = -p\log p - q\log q$

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) = ?$
5. $X \sim (p, q)$, $p + q = 1$
   - $H(X) = H(p,q) = -p\log p - q\log q$
   - $H(1,0) = (0,1) = 0$

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) = ?$
5. $X \sim (p, q)$, $p + q = 1$
   - $H(X) = H(p,q) = -p\log p - q\log q$
   - $H(1,0) = (0,1) = 0$
   - $H(\frac{1}{2}, \frac{1}{2}) = 1$

Examples

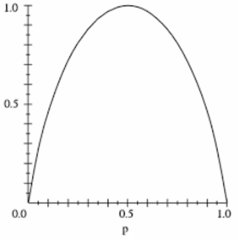
1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) =?$
5. $X \sim (p, q)$, $p + q = 1$
   - $H(X) = H(p,q) = -p\log p - q\log q$
   - $H(1,0) = (0,1) = 0$
   - $H(\frac{1}{2}, \frac{1}{2}) = 1$
   - $h(p) := H(p, 1-p)$ is continuous

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathrm{P}_X(x_1) = \frac{1}{2}, \mathrm{P}_X(x_2) = \frac{1}{4}, \mathrm{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathrm{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) =?$
5. $X \sim (p, q)$, $p + q = 1$
   - $H(X) = H(p,q) = -p\log p - q\log q$
   - $H(1,0) = (0,1) = 0$
   - $H(\frac{1}{2}, \frac{1}{2}) = 1$
   - $h(p) := H(p, 1-p)$ is continuous

Examples

1. $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$:
   (i.e., for some $x_1 \neq x_2 \neq x_3$, $\mathsf{P}_X(x_1) = \frac{1}{2}, \mathsf{P}_X(x_2) = \frac{1}{4}, \mathsf{P}_X(x_3) = \frac{1}{4}$)
   $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1\frac{1}{2}$.
2. $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.
3. X is uniformly distributed over $\{0,1\}^n$:
   $H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n}\log\frac{1}{2^n} = -\log\frac{1}{2^n} = n$.
   - n bits are needed to describe X
   - n bits are needed to sample X
4. $X = X_1, \ldots, X_n$ where X_i are iid over $\{0,1\}$, with $\mathsf{P}_{X_i}(1) := \Pr[X_i = 1] = \frac{1}{3}$. $H(X) = ?$
5. $X \sim (p, q)$, $p + q = 1$
   - $H(X) = H(p,q) = -p\log p - q\log q$
   - $H(1,0) = (0,1) = 0$
   - $H(\frac{1}{2}, \frac{1}{2}) = 1$
   - $h(p) := H(p, 1-p)$ is continuous



Infinite random variables

- For infinite (discrete) random variable $X \sim (p_1, p_2, \ldots)$, $H(X) = \sum_{i=1}^{\infty} p_i \log p_i$

Infinite random variables

- For infinite (discrete) random variable $X \sim (p_1, p_2, \ldots)$, $H(X) = \sum_{i=1}^{\infty} p_i \log p_i$
- Might be finite: $p_i = 2^{-i}$

Infinite random variables

- For infinite (discrete) random variable $X \sim (p_1, p_2, \ldots)$, $H(X) = \sum_{i=1}^{\infty} p_i \log p_i$
- Might be finite: $p_i = 2^{-i}$

  Hence, $H(X) = \sum_{i=1}^{\infty} 2^{-i} \cdot i < \infty$

Infinite random variables

- For infinite (discrete) random variable $X \sim (p_1, p_2, \ldots)$, $H(X) = \sum_{i=1}^{\infty} p_i \log p_i$
- Might be finite: $p_i = 2^{-i}$

  Hence, $H(X) = \sum_{i=1}^{\infty} 2^{-i} \cdot i < \infty$
- Or infinite, $X \sim \{p_{i,j}\}_{i \in \mathbb{N}, j \in \{1,\ldots,2^{2^i - i}\}}$, for $p_{i,j} = 2^{-2^i}$.

Infinite random variables

- For infinite (discrete) random variable $X \sim (p_1, p_2, \ldots)$, $H(X) = \sum_{i=1}^{\infty} p_i \log p_i$
- Might be finite: $p_i = 2^{-i}$

  Hence, $H(X) = \sum_{i=1}^{\infty} 2^{-i} \cdot i < \infty$
- Or infinite, $X \sim \{p_{i,j}\}_{i \in \mathbb{N}, j \in \{1,\ldots,2^{2^i - i}\}}$, for $p_{i,j} = 2^{-2^i}$.

  Hence, $H(X) = \sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i - i}} 2^{-2^i} \cdot 2^i = \sum_{i=1}^{\infty} 2^{-i} \cdot 2^i = \infty$

Infinite random variables

- For infinite (discrete) random variable $X \sim (p_1, p_2, \ldots)$, $H(X) = \sum_{i=1}^{\infty} p_i \log p_i$
- Might be finite: $p_i = 2^{-i}$

  Hence, $H(X) = \sum_{i=1}^{\infty} 2^{-i} \cdot i < \infty$
- Or infinite, $X \sim \{p_{i,j}\}_{i \in \mathbb{N}, j \in \{1,\ldots,2^{2^i - i}\}}$, for $p_{i,j} = 2^{-2^i}$.

  Hence, $H(X) = \sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i - i}} 2^{-2^i} \cdot 2^i = \sum_{i=1}^{\infty} 2^{-i} \cdot 2^i = \infty$
- We will typically restrict our attention to finite random variables.

Applications

- Data compression

Applications

- Data compression
- Error correction codes

Applications

- Data compression
- Error correction codes
- Algorithm Analysis

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube
  - Projection of Q on xy — 6

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube
  - Projection of Q on xy — 6
  - Projection of Q on xz — 8

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube
  - Projection of Q on xy — 6
  - Projection of Q on xz — 8
  - Projection of Q on yz — 12

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube
  - Projection of Q on xy — 6
  - Projection of Q on xz — 8
  - Projection of Q on yz — 12

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube
  - Projection of Q on xy — 6
  - Projection of Q on xz — 8
  - Projection of Q on yz — 12

  Can we bound $|Q|$?

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube
  - Projection of Q on xy — 6
  - Projection of Q on xz — 8
  - Projection of Q on yz — 12

  Can we bound $|Q|$?
- and more and more...

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube
  - Projection of Q on xy — 6
  - Projection of Q on xz — 8
  - Projection of Q on yz — 12

  Can we bound $|Q|$?
- and more and more...

Applications

- Data compression
- Error correction codes
- Algorithm Analysis
- Protocols Analysis
- Cryptography
- Counting. Example # of gold coins in a cube
  - Projection of Q on xy — 6
  - Projection of Q on xz — 8
  - Projection of Q on yz — 12

  Can we bound $|Q|$?
- and more and more...

And all are rather simple to prove

Axiomatic derivation of the entropy function

Any other choices for defining entropy?

Axiomatic derivation of the entropy function

Any other choices for defining entropy?
Shannon function is the only symmetric function (over probability distributions) satisfying the following three axioms:

A1 Continuity: $H(p, 1-p)$ is continuous function of p.

A2 Normalization: $H(\frac{1}{2}, \frac{1}{2}) = 1$

A3 Grouping axiom:
$H(p_1, p_2, \ldots, p_m) = H(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

Axiomatic derivation of the entropy function

Any other choices for defining entropy?
Shannon function is the only symmetric function (over probability distributions) satisfying the following three axioms:

A1 Continuity: $H(p, 1-p)$ is continuous function of p.

A2 Normalization: $H(\frac{1}{2}, \frac{1}{2}) = 1$

A3 Grouping axiom:
$H(p_1, p_2, \ldots, p_m) = H(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

Why $A3$?

Axiomatic derivation of the entropy function

Any other choices for defining entropy?
Shannon function is the only symmetric function (over probability distributions) satisfying the following three axioms:

- **A1** Continuity: $H(p, 1-p)$ is continuous function of p.
- **A2** Normalization: $H(\frac{1}{2}, \frac{1}{2}) = 1$
- **A3** Grouping axiom:
  $H(p_1, p_2, \ldots, p_m) = H(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

Why $A3$?

Not hard to prove that Shannon's entropy function satisfies above axioms, proving this is the only such function is more challenging.

Axiomatic derivation of the entropy function

Any other choices for defining entropy?
Shannon function is the only symmetric function (over probability distributions) satisfying the following three axioms:

- **A1** Continuity: $H(p, 1-p)$ is continuous function of p.
- **A2** Normalization: $H(\frac{1}{2}, \frac{1}{2}) = 1$
- **A3** Grouping axiom:
  $H(p_1, p_2, \ldots, p_m) = H(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

Why $A3$?

Not hard to prove that Shannon's entropy function satisfies above axioms, proving this is the only such function is more challenging.

Let H^* be a function that satisfying the above axioms.

Axiomatic derivation of the entropy function

Any other choices for defining entropy?
Shannon function is the only symmetric function (over probability distributions) satisfying the following three axioms:

A1 Continuity: $H(p, 1-p)$ is continuous function of p.

A2 Normalization: $H(\frac{1}{2}, \frac{1}{2}) = 1$

A3 Grouping axiom:
$H(p_1, p_2, \ldots, p_m) = H(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

Why $A3$?

Not hard to prove that Shannon's entropy function satisfies above axioms, proving this is the only such function is more challenging.

Let H^* be a function that satisfying the above axioms.

We prove (assuming additional axiom) that H^* is the Shannon function H.

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.

Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.
Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.
Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Proof:

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.
Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Proof: Let $h(q) = H^*(q, 1-q)$.

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.
Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Proof: Let $h(q) = H^*(q, 1-q)$.

$$H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 h(\frac{p_2}{S_2}) \quad (1)$$

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.
Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Proof: Let $h(q) = H^*(q, 1-q)$.

$$\begin{aligned} H^*(p_1, p_2, \ldots, p_m) &= H^*(S_2, p_3, \ldots, p_m) + S_2 h(\frac{p_2}{S_2}) \qquad (1) \\ &= H^*(S_3, p_4, \ldots, p_m) + S_3 h(\frac{p_3}{S_3}) + S_2 h(\frac{p_2}{S_2}) \end{aligned}$$

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.
Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Proof: Let $h(q) = H^*(q, 1-q)$.

$$\begin{aligned} H^*(p_1, p_2, \ldots, p_m) &= H^*(S_2, p_3, \ldots, p_m) + S_2 h(\frac{p_2}{S_2}) \qquad (1) \\ &= H^*(S_3, p_4, \ldots, p_m) + S_3 h(\frac{p_3}{S_3}) + S_2 h(\frac{p_2}{S_2}) \\ &\vdots \\ &= H^*(S_k, p_{k+1}, \ldots, p_m) + \sum_{i=2}^{k} S_i h(\frac{p_i}{S_i}) \end{aligned}$$

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.
Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Proof: Let $h(q) = H^*(q, 1-q)$.

$$\begin{aligned} H^*(p_1, p_2, \ldots, p_m) &= H^*(S_2, p_3, \ldots, p_m) + S_2 h(\frac{p_2}{S_2}) \qquad (1) \\ &= H^*(S_3, p_4, \ldots, p_m) + S_3 h(\frac{p_3}{S_3}) + S_2 h(\frac{p_2}{S_2}) \\ &\vdots \\ &= H^*(S_k, p_{k+1}, \ldots, p_m) + \sum_{i=2}^{k} S_i h(\frac{p_i}{S_i}) \end{aligned}$$

Hence,

$$H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k}) = H^*(\frac{S_{k-1}}{S_k}, \frac{p_k}{S_k}) + \sum_{i=2}^{k-1} \frac{S_i}{S_k} h(\frac{p_i/S_k}{S_i/S_k})$$

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.

Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Proof: Let $h(q) = H^*(q, 1-q)$.

$$\begin{aligned} H^*(p_1, p_2, \ldots, p_m) &= H^*(S_2, p_3, \ldots, p_m) + S_2 h(\frac{p_2}{S_2}) \\ &= H^*(S_3, p_4, \ldots, p_m) + S_3 h(\frac{p_3}{S_3}) + S_2 h(\frac{p_2}{S_2}) \\ &\vdots \\ &= H^*(S_k, p_{k+1}, \ldots, p_m) + \sum_{i=2}^{k} S_i h(\frac{p_i}{S_i}) \end{aligned} \tag{1}$$

Hence,

$$H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k}) = H^*(\frac{S_{k-1}}{S_k}, \frac{p_k}{S_k}) + \sum_{i=2}^{k-1} \frac{S_i}{S_k} h(\frac{p_i/S_k}{S_i/S_k}) = \frac{1}{S_k} \sum_{i=2}^{k} S_i h(\frac{p_i}{S_i}) \tag{2}$$

Generalization of the grouping axiom

Fix $p = (p_1, \ldots, p_m)$ and let $S_k = \sum_{i=1}^{k} p_i$.
Grouping axiom: $H^*(p_1, p_2, \ldots, p_m) = H^*(S_2, p_3, \ldots, p_m) + S_2 H^*(\frac{p_1}{S_2}, \frac{p_2}{S_2})$.

Claim 1 (Generalized grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) = H^*(S_k, p_{k+1}, \ldots, p_m) + S_k \cdot H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k})$

Proof: Let $h(q) = H^*(q, 1-q)$.

$$\begin{aligned} H^*(p_1, p_2, \ldots, p_m) &= H^*(S_2, p_3, \ldots, p_m) + S_2 h(\frac{p_2}{S_2}) \qquad (1) \\ &= H^*(S_3, p_4, \ldots, p_m) + S_3 h(\frac{p_3}{S_3}) + S_2 h(\frac{p_2}{S_2}) \\ &\vdots \\ &= H^*(S_k, p_{k+1}, \ldots, p_m) + \sum_{i=2}^{k} S_i h(\frac{p_i}{S_i}) \end{aligned}$$

Hence,

$$H^*(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k}) = H^*(\frac{S_{k-1}}{S_k}, \frac{p_k}{S_k}) + \sum_{i=2}^{k-1} \frac{S_i}{S_k} h(\frac{p_i/S_k}{S_i/S_k}) = \frac{1}{S_k} \sum_{i=2}^{k} S_i h(\frac{p_i}{S_i}) \qquad (2)$$

Claim follows by combining the above equations. □

Further generalization of the grouping axiom

Let $1 = k_1 < k_2 < \ldots < k_q < m$ and let $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$ (letting $k_{q+1} = m+1$).

Claim 2 (Generalized^{++} grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) =$
$H^*(C_1, \ldots, C_q) + C_1 \cdot H^*(\frac{p_1}{C_1}, \ldots, \frac{p_{k_2-1}}{C_1}) + \ldots + C_q \cdot H^*(\frac{p_{k_q+1}}{C_q}, \ldots, \frac{p_m}{C_q})$

Further generalization of the grouping axiom

Let $1 = k_1 < k_2 < \ldots < k_q < m$ and let $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$ (letting $k_{q+1} = m+1$).

Claim 2 (Generalized^{++} grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) =$
$H^*(C_1, \ldots, C_q) + C_1 \cdot H^*(\frac{p_1}{C_1}, \ldots, \frac{p_{k_2-1}}{C_1}) + \ldots + C_q \cdot H^*(\frac{p_{k_q+1}}{C_q}, \ldots, \frac{p_m}{C_q})$

Proof:

Further generalization of the grouping axiom

Let $1 = k_1 < k_2 < \ldots < k_q < m$ and let $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$ (letting $k_{q+1} = m + 1$).

Claim 2 (Generalized^{++} grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) =$
$H^*(C_1, \ldots, C_q) + C_1 \cdot H^*(\frac{p_1}{C_1}, \ldots, \frac{p_{k_2-1}}{C_1}) + \ldots + C_q \cdot H^*(\frac{p_{k_q+1}}{C_q}, \ldots, \frac{p_m}{C_q})$

Proof: Follow by the extended group axiom and the symmetry of H $\square$

Further generalization of the grouping axiom

Let $1 = k_1 < k_2 < \ldots < k_q < m$ and let $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$ (letting $k_{q+1} = m+1$).

Claim 2 (Generalized^{++} grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) =$
$H^*(C_1, \ldots, C_q) + C_1 \cdot H^*(\frac{p_1}{C_1}, \ldots, \frac{p_{k_2-1}}{C_1}) + \ldots + C_q \cdot H^*(\frac{p_{k_q+1}}{C_q}, \ldots, \frac{p_m}{C_q})$

Proof: Follow by the extended group axiom and the symmetry of H $\Box$

Implication: Let $f(m) := H^*(\underbrace{\frac{1}{m}, \ldots, \frac{1}{m}}_{m})$

Further generalization of the grouping axiom

Let $1 = k_1 < k_2 < \ldots < k_q < m$ and let $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$ (letting $k_{q+1} = m + 1$).

Claim 2 (Generalized^{++} grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) =$
$H^*(C_1, \ldots, C_q) + C_1 \cdot H^*(\frac{p_1}{C_1}, \ldots, \frac{p_{k_2-1}}{C_1}) + \ldots + C_q \cdot H^*(\frac{p_{k_q+1}}{C_q}, \ldots, \frac{p_m}{C_q})$

Proof: Follow by the extended group axiom and the symmetry of H $\square$

Implication: Let $f(m) := H^*(\underbrace{\frac{1}{m}, \ldots, \frac{1}{m}}_{m})$

- $f(3^2) = 2f(3) = 2H^*(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

Further generalization of the grouping axiom

Let $1 = k_1 < k_2 < \ldots < k_q < m$ and let $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$ (letting $k_{q+1} = m+1$).

Claim 2 (Generalized^{++} grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) =$
$H^*(C_1, \ldots, C_q) + C_1 \cdot H^*(\frac{p_1}{C_1}, \ldots, \frac{p_{k_2-1}}{C_1}) + \ldots + C_q \cdot H^*(\frac{p_{k_q+1}}{C_q}, \ldots, \frac{p_m}{C_q})$

Proof: Follow by the extended group axiom and the symmetry of H $\square$

Implication: Let $f(m) := H^*(\underbrace{\frac{1}{m}, \ldots, \frac{1}{m}}_{m})$

- $f(3^2) = 2f(3) = 2H^*(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
  $\implies f(3^n) = nf(3)$.

Further generalization of the grouping axiom

Let $1 = k_1 < k_2 < \ldots < k_q < m$ and let $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$ (letting $k_{q+1} = m+1$).

Claim 2 (Generalized^{++} grouping axiom)

$H^*(p_1, p_2, \ldots, p_m) =$
$H^*(C_1, \ldots, C_q) + C_1 \cdot H^*(\frac{p_1}{C_1}, \ldots, \frac{p_{k_2-1}}{C_1}) + \ldots + C_q \cdot H^*(\frac{p_{k_q+1}}{C_q}, \ldots, \frac{p_m}{C_q})$

Proof: Follow by the extended group axiom and the symmetry of H $\square$

Implication: Let $f(m) := H^*(\underbrace{\frac{1}{m}, \ldots, \frac{1}{m}}_{m})$

- $f(3^2) = 2f(3) = 2H^*(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
  $\implies f(3^n) = nf(3)$.
- $f(mn) = f(m) + f(n)$

Further generalization of the grouping axiom

Let $1 = k_1 < k_2 < \ldots < k_q < m$ and let $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$ (letting $k_{q+1} = m+1$).

Claim 2 (Generalized^{++} grouping axiom)

$$H^*(p_1, p_2, \ldots, p_m) =$$
$$H^*(C_1, \ldots, C_q) + C_1 \cdot H^*(\tfrac{p_1}{C_1}, \ldots, \tfrac{p_{k_2-1}}{C_1}) + \ldots + C_q \cdot H^*(\tfrac{p_{k_q+1}}{C_q}, \ldots, \tfrac{p_m}{C_q})$$

Proof: Follow by the extended group axiom and the symmetry of H $\square$

Implication: Let $f(m) := H^*(\underbrace{\frac{1}{m}, \ldots, \frac{1}{m}}_{m})$

- $f(3^2) = 2f(3) = 2H^*(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

  $\implies f(3^n) = nf(3)$.
- $f(mn) = f(m) + f(n)$

  $\implies f(m^k) = kf(m)$

$f(m) = \log m$

$f(m) = \log m$

We give a proof under the additional axiom

A4 $f(m) \leq f(m+1)$

(you can Google for a proof using only $A1$–$A3$)

$f(m) = \log m$

We give a proof under the additional axiom

A4 $f(m) \leq f(m+1)$

(you can Google for a proof using only $A1$–$A3$)

- For $n \in \mathbb{N}$, let $k = \lfloor \log 3^n = n \log 3 \rfloor$.

$f(m) = \log m$

We give a proof under the additional axiom

A4 $f(m) \leq f(m+1)$

(you can Google for a proof using only $A1$–$A3$)

- For $n \in \mathbb{N}$, let $k = \lfloor \log 3^n = n \log 3 \rfloor$.
- Since, $2^k \leq 3^n \leq 2^{k+1}$, by $A4$: $f(2^k) \leq f(3^n) \leq f(2^{k+1})$.

$f(m) = \log m$

We give a proof under the additional axiom

A4 $f(m) \leq f(m+1)$

(you can Google for a proof using only $A1$–$A3$)

- For $n \in \mathbb{N}$, let $k = \lfloor \log 3^n = n \log 3 \rfloor$.
- Since, $2^k \leq 3^n \leq 2^{k+1}$, by $A4$: $\quad f(2^k) \leq f(3^n) \leq f(2^{k+1})$.
- By grouping axiom, $k < nf(3) < k + 1$.

$f(m) = \log m$

We give a proof under the additional axiom

A4 $f(m) \leq f(m+1)$

(you can Google for a proof using only $A1$–$A3$)

- For $n \in \mathbb{N}$, let $k = \lfloor \log 3^n = n \log 3 \rfloor$.
- Since, $2^k \leq 3^n \leq 2^{k+1}$, by $A4$: $f(2^k) \leq f(3^n) \leq f(2^{k+1})$.
- By grouping axiom, $k < nf(3) < k+1$.

$\implies \frac{\lfloor n \log 3 \rfloor}{n} \leq f(3) \leq \frac{\lfloor n \log 3 \rfloor + 1}{n}$ for any $n \in \mathbb{N}$

$f(m) = \log m$

We give a proof under the additional axiom

A4 $f(m) \leq f(m+1)$

(you can Google for a proof using only $A1$–$A3$)

- For $n \in \mathbb{N}$, let $k = \lfloor \log 3^n = n \log 3 \rfloor$.
- Since, $2^k \leq 3^n \leq 2^{k+1}$, by $A4$: $f(2^k) \leq f(3^n) \leq f(2^{k+1})$.
- By grouping axiom, $k < nf(3) < k+1$.

$\implies \frac{\lfloor n \log 3 \rfloor}{n} \leq f(3) \leq \frac{\lfloor n \log 3 \rfloor + 1}{n}$ for any $n \in \mathbb{N}$

$\implies f(3) = \log 3$.

$f(m) = \log m$

We give a proof under the additional axiom

A4 $f(m) \leq f(m+1)$

(you can Google for a proof using only $A1$–$A3$)

- For $n \in \mathbb{N}$, let $k = \lfloor \log 3^n = n \log 3 \rfloor$.
- Since, $2^k \leq 3^n \leq 2^{k+1}$, by $A4$: $f(2^k) \leq f(3^n) \leq f(2^{k+1})$.
- By grouping axiom, $k < nf(3) < k+1$.

$\Longrightarrow$ $\frac{\lfloor n \log 3 \rfloor}{n} \leq f(3) \leq \frac{\lfloor n \log 3 \rfloor + 1}{n}$ for any $n \in \mathbb{N}$

$\Longrightarrow$ $f(3) = \log 3$.

- Proof extends to any integer (not only 3)

$H^*(p, q) = -p \log p - q \log q$

$H^*(p, q) = -p \log p - q \log q$

- For rational p, q, let $p = \frac{k}{m}$ and $q = \frac{m-k}{m}$, where m is the smallest common multiplier.

$H^*(p,q) = -p\log p - q\log q$

- For rational p, q, let $p = \frac{k}{m}$ and $q = \frac{m-k}{m}$, where m is the smallest common multiplier.

$H^*(p,q) = -p\log p - q\log q$

- For rational p,q, let $p = \frac{k}{m}$ and $q = \frac{m-k}{m}$, where m is the smallest common multiplier.
- By grouping axiom, $f(m) = H^*(p,q) + p \cdot f(k) + q \cdot f(m-k)$.

$H^*(p,q) = -p\log p - q\log q$

- For rational p, q, let $p = \frac{k}{m}$ and $q = \frac{m-k}{m}$, where m is the smallest common multiplier.
- By grouping axiom, $f(m) = H^*(p,q) + p \cdot f(k) + q \cdot f(m-k)$.
- Hence,

$$H^*(p,q) \quad = \quad \log m - p\log k - q\log(m-k)$$

$H^*(p,q) = -p\log p - q\log q$

- For rational p, q, let $p = \frac{k}{m}$ and $q = \frac{m-k}{m}$, where m is the smallest common multiplier.
- By grouping axiom, $f(m) = H^*(p,q) + p \cdot f(k) + q \cdot f(m-k)$.
- Hence,

$$H^*(p,q) = \log m - p\log k - q\log(m-k)$$

$H^*(p,q) = -p\log p - q\log q$

- For rational p,q, let $p = \frac{k}{m}$ and $q = \frac{m-k}{m}$, where m is the smallest common multiplier.
- By grouping axiom, $f(m) = H^*(p,q) + p\cdot f(k) + q\cdot f(m-k)$.
- Hence,

$$
\begin{aligned}
H^*(p,q) &= \log m - p\log k - q\log(m-k)\\
&= p(\log m - \log k) + q(\log m - \log(m-k))
\end{aligned}
$$

$H^*(p, q) = -p \log p - q \log q$

- For rational p, q, let $p = \frac{k}{m}$ and $q = \frac{m-k}{m}$, where m is the smallest common multiplier.
- By grouping axiom, $f(m) = H^*(p, q) + p \cdot f(k) + q \cdot f(m-k)$.
- Hence,

$$\begin{aligned} H^*(p, q) &= \log m - p \log k - q \log(m-k) \\ &= p(\log m - \log k) + q(\log m - \log(m-k)) \\ &= -p \log \frac{m}{k} - q \log \frac{m-k}{m} = -p \log p - q \log q \end{aligned}$$

$H^*(p,q) = -p\log p - q\log q$

- For rational p, q, let $p = \frac{k}{m}$ and $q = \frac{m-k}{m}$, where m is the smallest common multiplier.
- By grouping axiom, $f(m) = H^*(p,q) + p \cdot f(k) + q \cdot f(m-k)$.
- Hence,

$$\begin{aligned} H^*(p,q) &= \log m - p\log k - q\log(m-k) \\ &= p(\log m - \log k) + q(\log m - \log(m-k)) \\ &= -p\log\frac{m}{k} - q\log\frac{m-k}{m} = -p\log p - q\log q \end{aligned}$$

- By continuity axiom, holds for every p, q.

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

- For rational p_1, p_2, p_3, let $p_1 = \frac{k_1}{m}$, $q = \frac{k_2}{m}$ and $p_3 = \frac{k_3}{m}$, where $m = k_1 + k_2 + k_3$ is the smallest common multiplier.

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

- For rational p_1, p_2, p_3, let $p_1 = \frac{k_1}{m}$, $q = \frac{k_2}{m}$ and $p_3 = \frac{k_3}{m}$, where $m = k_1 + k_2 + k_3$ is the smallest common multiplier.

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

- For rational p_1, p_2, p_3, let $p_1 = \frac{k_1}{m}$, $q = \frac{k_2}{m}$ and $p_3 = \frac{k_3}{m}$, where $m = k_1 + k_2 + k_3$ is the smallest common multiplier.
- $f(m) = H^*(p_1, p_2, p_3) + p_1 f(k_1) + p_2 f(k_2) + p_3 f(k_3)$

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

- For rational p_1, p_2, p_3, let $p_1 = \frac{k_1}{m}$, $q = \frac{k_2}{m}$ and $p_3 = \frac{k_3}{m}$, where $m = k_1 + k_2 + k_3$ is the smallest common multiplier.
- $f(m) = H^*(p_1, p_2, p_3) + p_1 f(k_1) + p_2 f(k_2) + p_3 f(k_3)$
- Hence,

$$H^*(p_1, p_2, p_3) \quad = \quad \log m - p_1 \log k_1 - p_2 \log k_2 - p_3 \log k_3$$

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

- For rational p_1, p_2, p_3, let $p_1 = \frac{k_1}{m}$, $q = \frac{k_2}{m}$ and $p_3 = \frac{k_3}{m}$, where $m = k_1 + k_2 + k_3$ is the smallest common multiplier.
- $f(m) = H^*(p_1, p_2, p_3) + p_1 f(k_1) + p_2 f(k_2) + p_3 f(k_3)$
- Hence,

$$H^*(p_1, p_2, p_3) \quad = \quad \log m - p_1 \log k_1 - p_2 \log k_2 - p_3 \log k_3$$

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

- For rational p_1, p_2, p_3, let $p_1 = \frac{k_1}{m}$, $q = \frac{k_2}{m}$ and $p_3 = \frac{k_3}{m}$, where $m = k_1 + k_2 + k_3$ is the smallest common multiplier.
- $f(m) = H^*(p_1, p_2, p_3) + p_1 f(k_1) + p_2 f(k_2) + p_3 f(k_3)$
- Hence,

$$\begin{aligned} H^*(p_1, p_2, p_3) &= \log m - p_1 \log k_1 - p_2 \log k_2 - p_3 \log k_3 \\ &= -p_1 \log \frac{k_1}{m} - p_2 \log \frac{k_2}{m} - p_3 \frac{k_3}{m} \end{aligned}$$

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

- For rational p_1, p_2, p_3, let $p_1 = \frac{k_1}{m}$, $q = \frac{k_2}{m}$ and $p_3 = \frac{k_3}{m}$, where $m = k_1 + k_2 + k_3$ is the smallest common multiplier.
- $f(m) = H^*(p_1, p_2, p_3) + p_1 f(k_1) + p_2 f(k_2) + p_3 f(k_3)$
- Hence,

$$
\begin{aligned}
H^*(p_1, p_2, p_3) &= \log m - p_1 \log k_1 - p_2 \log k_2 - p_3 \log k_3 \\
&= -p_1 \log \frac{k_1}{m} - p_2 \log \frac{k_2}{m} - p_3 \frac{k_3}{m} \\
&= -p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3
\end{aligned}
$$

$H^*(p_1, p_2, \ldots, p_m) = -\sum_i^m p_i \log p_i$

We prove for $m = 3$. Proof for arbitrary m follows the same lines.

- For rational p_1, p_2, p_3, let $p_1 = \frac{k_1}{m}$, $q = \frac{k_2}{m}$ and $p_3 = \frac{k_3}{m}$, where $m = k_1 + k_2 + k_3$ is the smallest common multiplier.
- $f(m) = H^*(p_1, p_2, p_3) + p_1 f(k_1) + p_2 f(k_2) + p_3 f(k_3)$
- Hence,

$$\begin{aligned} H^*(p_1, p_2, p_3) &= \log m - p_1 \log k_1 - p_2 \log k_2 - p_3 \log k_3 \\ &= -p_1 \log \frac{k_1}{m} - p_2 \log \frac{k_2}{m} - p_3 \frac{k_3}{m} \\ &= -p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3 \end{aligned}$$

- By continuity axiom, holds for every p_1, p_2, p_3.

□

Section 1

Basic Properties

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
- Non negativity is clear.

$0 \le H(p_1, \ldots, p_m) \le \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
- Non negativity is clear.
- A function f is concave ("keura") if $\forall\ t_1, t_2$, $\lambda \in [0, 1] \le 1$
  $\lambda f(t_1) + (1 - \lambda) f(t_2) \le f(\lambda t_1 + (1 - \lambda) t_2)$

$0 \le H(p_1, \ldots, p_m) \le \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
- Non negativity is clear.
- A function f is concave ("keura") if $\forall\ t_1, t_2$, $\lambda \in [0,1] \le 1$
  $\lambda f(t_1) + (1-\lambda) f(t_2) \le f(\lambda t_1 + (1-\lambda) t_2)$

$\implies$ (by induction) $\forall\ t_1, \ldots, t_k$, $\lambda_1, \ldots, \lambda_k \in [0,1]$ with $\sum_i \lambda_i = 1$
$\sum_i \lambda_i f(t_i) \le f(\sum_i \lambda_i t_i)$

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
- Non negativity is clear.
- A function f is concave ("keura") if $\forall\ t_1, t_2$, $\lambda \in [0, 1] \leq 1$
  $\lambda f(t_1) + (1 - \lambda) f(t_2) \leq f(\lambda t_1 + (1 - \lambda) t_2)$

$\implies$ (by induction) $\forall\ t_1, \ldots, t_k$, $\lambda_1, \ldots, \lambda_k \in [0, 1]$ with $\sum_i \lambda_i = 1$
$\sum_i \lambda_i f(t_i) \leq f(\sum_i \lambda_i t_i)$

$\implies$ (Jensen inequality): $\mathrm{E}\, f(X) \leq f(\mathrm{E}\, X)$ for any random variable X.

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
- Non negativity is clear.
- A function f is concave ("keura") if $\forall\ t_1, t_2$, $\lambda \in [0, 1] \leq 1$
  $\lambda f(t_1) + (1 - \lambda) f(t_2) \leq f(\lambda t_1 + (1 - \lambda) t_2)$

$\implies$ (by induction) $\forall\ t_1, \ldots, t_k$, $\lambda_1, \ldots, \lambda_k \in [0, 1]$ with $\sum_i \lambda_i = 1$
$\sum_i \lambda_i f(t_i) \leq f(\sum_i \lambda_i t_i)$

$\implies$ (Jensen inequality): $\mathrm{E}\, f(X) \leq f(\mathrm{E}\, X)$ for any random variable X.

- $\log(x)$ is (strictly) concave for $x > 0$, since its second derivative $(-\frac{1}{x^2})$ is always negative.

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
- Non negativity is clear.
- A function f is concave ("keura") if $\forall\ t_1, t_2$, $\lambda \in [0, 1] \leq 1$
  $\lambda f(t_1) + (1 - \lambda) f(t_2) \leq f(\lambda t_1 + (1 - \lambda) t_2)$

$\Longrightarrow$ (by induction) $\forall\ t_1, \ldots, t_k$, $\lambda_1, \ldots, \lambda_k \in [0, 1]$ with $\sum_i \lambda_i = 1$
$\sum_i \lambda_i f(t_i) \leq f(\sum_i \lambda_i t_i)$

$\Longrightarrow$ (Jensen inequality): $\mathrm{E}\, f(X) \leq f(\mathrm{E}\, X)$ for any random variable X.

- $\log(x)$ is (strictly) concave for $x > 0$, since its second derivative $(-\frac{1}{x^2})$ is always negative.
- Hence, $H(p_1, \ldots, p_m) = \sum_i p_i \log \frac{1}{p_i} \leq \log \sum_i p_i \frac{1}{p_i} = \log m$

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
- Non negativity is clear.
- A function f is concave ("keura") if $\forall\ t_1, t_2, \lambda \in [0, 1] \leq 1$
  $\lambda f(t_1) + (1 - \lambda) f(t_2) \leq f(\lambda t_1 + (1 - \lambda) t_2)$

$\implies$ (by induction) $\forall\ t_1, \ldots, t_k$, $\lambda_1, \ldots, \lambda_k \in [0, 1]$ with $\sum_i \lambda_i = 1$
$\sum_i \lambda_i f(t_i) \leq f(\sum_i \lambda_i t_i)$

$\implies$ (Jensen inequality): $\mathrm{E}\, f(X) \leq f(\mathrm{E}\, X)$ for any random variable X.

- $\log(x)$ is (strictly) concave for $x > 0$, since its second derivative $(-\frac{1}{x^2})$ is always negative.
- Hence, $H(p_1, \ldots, p_m) = \sum_i p_i \log \frac{1}{p_i} \leq \log \sum_i p_i \frac{1}{p_i} = \log m$
- Alternatively, for X over $\{1, \ldots, m\}$,
  $H(X) = \mathrm{E}_X \log \frac{1}{\mathrm{P}_X(X)} \leq \log \mathrm{E}_X \frac{1}{\mathrm{P}_X(X)} = \log m$

$0 \leq H(p_1, \ldots, p_m) \leq \log m$

- Tight bounds
  - $H(p_1, \ldots, p_m) = 0$ for $(p_1, \ldots, p_m) = (1, 0, \ldots, 0)$.
  - $H(p_1, \ldots, p_m) = \log m$ for $(p_1, \ldots, p_m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
- Non negativity is clear.
- A function f is concave ("keura") if $\forall\ t_1, t_2,\ \lambda \in [0, 1] \leq 1$ $\lambda f(t_1) + (1 - \lambda) f(t_2) \leq f(\lambda t_1 + (1 - \lambda) t_2)$

$\implies$ (by induction) $\forall\ t_1, \ldots, t_k$, $\lambda_1, \ldots, \lambda_k \in [0, 1]$ with $\sum_i \lambda_i = 1$ $\sum_i \lambda_i f(t_i) \leq f(\sum_i \lambda_i t_i)$

$\implies$ (Jensen inequality): $\mathrm{E}\, f(X) \leq f(\mathrm{E}\, X)$ for any random variable X.

- $\log(x)$ is (strictly) concave for $x > 0$, since its second derivative $(-\frac{1}{x^2})$ is always negative.
- Hence, $H(p_1, \ldots, p_m) = \sum_i p_i \log \frac{1}{p_i} \leq \log \sum_i p_i \frac{1}{p_i} = \log m$
- Alternatively, for X over $\{1, \ldots, m\}$, $H(X) = \mathrm{E}_X \log \frac{1}{\mathrm{P}_X(X)} \leq \log \mathrm{E}_X \frac{1}{\mathrm{P}_X(X)} = \log m$
- What if $\mathrm{Supp}(X) := \{x \colon \mathrm{P}_X(x) > 0\} \subsetneq [m]$?

$H(g(X)) \leq H(X)$

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\text{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.
  Proof:

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$H(X) = -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x)$$

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\text{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$\begin{aligned} H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x:\ g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x:\ g(x)=y} \log \mathrm{P}_X(x) \end{aligned}$$

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$
\begin{aligned}
H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) =
\end{aligned}
$$

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$
\begin{aligned}
H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y)
\end{aligned}
$$

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$
\begin{aligned}
H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x:\, g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x:\, g(x)=y} \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y)
\end{aligned}
$$

- Or use the group axiom...

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$\begin{aligned} H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y) \end{aligned}$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$\begin{aligned} H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y) \end{aligned}$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

  Proof:

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$\begin{aligned} H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x:\, g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x:\, g(x)=y} \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y) \end{aligned}$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

  Proof:

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$\begin{aligned} H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x:\, g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x:\, g(x)=y} \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y) \end{aligned}$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

  Proof: $p_X(X) = P_Y(Y)$.

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$\begin{aligned} H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y) \end{aligned}$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

  Proof: $p_X(X) = P_Y(Y)$.
- If g is non-injective (over $\mathrm{Supp}(X)$), then $H(Y) < H(X)$.

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$
\begin{aligned}
H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y)
\end{aligned}
$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

  Proof: $p_X(X) = P_Y(Y)$.
- If g is non-injective (over $\mathrm{Supp}(X)$), then $H(Y) < H(X)$.

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$\begin{aligned} H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\ &\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y) \end{aligned}$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

  Proof: $p_X(X) = P_Y(Y)$.
- If g is non-injective (over $\mathrm{Supp}(X)$), then $H(Y) < H(X)$. Proof: ?

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$
\begin{aligned}
H(X) &= -\sum_x \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_y \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\
&\geq -\sum_y \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y)
\end{aligned}
$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

  Proof: $p_X(X) = P_Y(Y)$.
- If g is non-injective (over $\mathrm{Supp}(X)$), then $H(Y) < H(X)$. Proof: ?
- $H(X) = H(2^X)$.

$H(g(X)) \leq H(X)$

Let X be a random variable, and let g be over $\mathrm{Supp}(X)$.

- $H(Y = g(X)) \leq H(X)$.

  Proof:

$$\begin{aligned} H(X) &= -\sum_{x} \mathrm{P}_X(x) \log \mathrm{P}_X(x) = -\sum_{y} \sum_{x\colon g(x)=y} \mathrm{P}_X(x) \log \mathrm{P}_X(x) \\ &\geq -\sum_{y} \mathrm{P}_Y(y) \cdot \max_{x\colon g(x)=y} \log \mathrm{P}_X(x) \\ &\geq -\sum_{y} \mathrm{P}_Y(y) \cdot \log \mathrm{P}_Y(y) = H(Y) \end{aligned}$$

- Or use the group axiom...
- If g is injective, then $H(Y) = H(X)$.

  Proof: $p_X(X) = P_Y(Y)$.
- If g is non-injective (over $\mathrm{Supp}(X)$), then $H(Y) < H(X)$. Proof: ?
- $H(X) = H(2^X)$.
- $H(\sin(X)) < H(X)$, if $0, \pi \in \mathrm{Supp}(X)$.

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$
- But the notion of entropy already existed in statistical physics

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$
- But the notion of entropy already existed in statistical physics
- There, entropy — energy that cannot used, statistical disorder

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$
- But the notion of entropy already existed in statistical physics
- There, entropy — energy that cannot used, statistical disorder
- Clausius (1865), who coined the name *entropy*, based on Carnot (1824), $H = \int_t \frac{\delta Q}{T} dt$ (Q is *heat* and T is *temperature*)

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$
- But the notion of entropy already existed in statistical physics
- There, entropy — energy that cannot used, statistical disorder
- Clausius (1865), who coined the name *entropy*, based on Carnot (1824), $H = \int_t \frac{\delta Q}{T} dt$ (Q is *heat* and T is *temperature*)
- Boltzmann (1877) $H = \log S$, for S being the number of states a system can be in (after measuring the macro parameters: pressure, temperature)

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$
- But the notion of entropy already existed in statistical physics
- There, entropy — energy that cannot used, statistical disorder
- Clausius (1865), who coined the name *entropy*, based on Carnot (1824), $H = \int_t \frac{\delta Q}{T} dt$ (Q is *heat* and T is *temperature*)
- Boltzmann (1877) $H = \log S$, for S being the number of states a system can be in (after measuring the macro parameters: pressure, temperature)
- $\log \#$ of states is Shannon entropy of the uniform distribution

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$
- But the notion of entropy already existed in statistical physics
- There, entropy — energy that cannot used, statistical disorder
- Clausius (1865), who coined the name *entropy*, based on Carnot (1824), $H = \int_t \frac{\delta Q}{T} dt$ (Q is *heat* and T is *temperature*)
- Boltzmann (1877) $H = \log S$, for S being the number of states a system can be in (after measuring the macro parameters: pressure, temperature)
- $\log \#$ of states is Shannon entropy of the uniform distribution
- Shannon looked for a name for his measure, von Neumann pointed out the relation to physics and suggested the name entropy.

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$
- But the notion of entropy already existed in statistical physics
- There, entropy — energy that cannot used, statistical disorder
- Clausius (1865), who coined the name *entropy*, based on Carnot (1824), $H = \int_t \frac{\delta Q}{T} dt$ (Q is *heat* and T is *temperature*)
- Boltzmann (1877) $H = \log S$, for S being the number of states a system can be in (after measuring the macro parameters: pressure, temperature)
- $\log \#$ of states is Shannon entropy of the uniform distribution
- Shannon looked for a name for his measure, von Neumann pointed out the relation to physics and suggested the name entropy.
- Today it is accepted that Shannon's entropy is the right notion also in statistical mechanic. Measures the uncertainty of a system — energy that cannot be used.

Historical background

- Shannon (1948) $H = -\sum_i p_i \log p_i$
- But the notion of entropy already existed in statistical physics
- There, entropy — energy that cannot used, statistical disorder
- Clausius (1865), who coined the name *entropy*, based on Carnot (1824), $H = \int_t \frac{\delta Q}{T} dt$ (Q is *heat* and T is *temperature*)
- Boltzmann (1877) $H = \log S$, for S being the number of states a system can be in (after measuring the macro parameters: pressure, temperature)
- $\log \ \#$ of states is Shannon entropy of the uniform distribution
- Shannon looked for a name for his measure, von Neumann pointed out the relation to physics and suggested the name entropy.
- Today it is accepted that Shannon's entropy is the right notion also in statistical mechanic. Measures the uncertainty of a system — energy that cannot be used.

- Carnot was also an engineer...

Notation

- $[n] = \{1, \ldots, n\}$

Notation

- $[n] = \{1, \ldots, n\}$
- $\mathrm{P}_X(x) = \Pr[X = x]$

Notation

- $[n] = \{1, \ldots, n\}$
- $\mathrm{P}_X(x) = \Pr[X = x]$
- $\mathrm{Supp}(X) := \{x \colon \mathrm{P}_X(x) > 0\}$

Notation

- $[n] = \{1, \ldots, n\}$
- $\mathrm{P}_X(x) = \Pr[X = x]$
- $\mathrm{Supp}(X) := \{x \colon \mathrm{P}_X(x) > 0\}$
- For random variable X over $\mathcal{X}$, let $p(x)$ be its density function: $p(x) = \mathrm{P}_X(x)$.

Notation

- $[n] = \{1, \ldots, n\}$
- $\mathrm{P}_X(x) = \Pr[X = x]$
- $\mathrm{Supp}(X) := \{x \colon \mathrm{P}_X(x) > 0\}$
- For random variable X over $\mathcal{X}$, let $p(x)$ be its density function: $p(x) = \mathrm{P}_X(x)$.

  In other words, $X \sim p(x)$.

Notation

- $[n] = \{1, \ldots, n\}$
- $P_X(x) = \Pr[X = x]$
- $\mathrm{Supp}(X) := \{x \colon P_X(x) > 0\}$
- For random variable X over $\mathcal{X}$, let $p(x)$ be its density function: $p(x) = P_X(x)$.

  In other words, $X \sim p(x)$.
- For random variable Y over $\mathcal{Y}$, let $p(y)$ be its density function: $p(y) = P_Y(y)\ldots$